基于改进UperNet的结球甘蓝叶球识别方法

朱轶萍[1,2], 吴华瑞[1,2,3,4*], 郭 旺[2,3,4], 吴小燕[2]

(1. 江苏大学 计算机科学与通信工程学院,江苏镇江 212013,中国; 2. 国家农业信息化工程技术研究中心,北京 100097,中国; 3. 北京市农林科学院信息技术研究中心,北京 100097,中国; 4. 农业农村部数字乡村技术重点实验室,北京 100097,中国)

摘 要: [**目的/意义**] 叶球是结球甘蓝的重要部分,其生长发育对田间管理至关重要。针对叶球分割识别存在大田背景复杂、光照不均匀和叶片纹理相似等问题,提出一种语义分割算法UperNet-ESA,旨在能快速、准确地分割田间场景中结球甘蓝的外叶和叶球,以实现田间结球甘蓝的智能化管理。[**方法**] 首先,采用统一感知解析网络(Unified Perceptual Parsing Network, UperNet)作为高效语义分割框架,将主干网络改为先进的ConvNeXt,使得模型在提升分割精度的同时也能具有较低的模型复杂度;其次,利用高效通道注意力机制(Efficient Channel Attention, ECA)融入特征提取网络的各阶段,进一步捕捉图像的细节信息;最后,通过将特征选择模块(Feature Selection Model, FSM)和特征对齐模块(Feature Alignment Model, FAM)集成到特征金字塔框架中,得到更为精确的目标边界预测结果。[**结果和讨论**] 在自制结球甘蓝图像数据集上进行实验,与目前主流的UNet、PSPNet和DeeplabV3+语义分割模型相比,改进UperNet方法的平均交并比为92.45%,平均像素准确率为94.32%,推理速度为16.6 f/s,能够达到最佳精度-速度平衡效果。[**结论**] 研究成果可为结球甘蓝生长智能化监测提供理论参考,对甘蓝产业发展具有重要的应用前景。

关键词: 结球甘蓝;语义分割;叶球识别;注意力机制;特征选择;特征对齐

中图分类号: TP183 **文献标志码:** A **文章编号:** SA202401020

引用格式:朱轶萍, 吴华瑞, 郭旺, 吴小燕. 基于改进UperNet的结球甘蓝叶球识别方法[J]. 智慧农业(中英文), 2024, 6(3): 128-137. DOI:10.12133/j.smartag.SA202401020

ZHU Yiping, WU Huarui, GUO Wang, WU Xiaoyan. Identification Method of Kale Leaf Ball Based on Improved UperNet [J]. Smart Agriculture, 2024, 6(3): 128-137. DOI:10.12133/j.smartag.SA202401020 (in Chinese with English abstract)

0 引 言

结球甘蓝(*Brassica oleracea* var. *capitata* L.),简称甘蓝,是世界上重要的大宗蔬菜品种。甘蓝定植后的生命周期主要分为莲座期和结球期,主要生长特征是其外叶和叶球,而甘蓝叶球性状可以指导田间水肥参数调整,以及影响甘蓝的最终产量[1]。在农业领域,快速精准全面地掌握作物的生长状态十分重要,利于及时采取有效的田间管理措施,能够减少甘蓝产业受损风险。目前甘蓝叶球区域的观察方式还是以人工为主,不仅费时费力,而且存在人员间偏差与实时性不高等客观问题。在大规模露地甘蓝种植模式下,利用田间自动巡检机器人对甘蓝生长状态进行监测,是未来人工智能技术在农业生产中应用[2]不断推广的趋势。

近年来,随着深度学习(Deep Learning, DL)[3,4]的发展和应用,越来越多的分割问题[5-7]由DL技术解决。通过使用大量标记数据集输入模型进行实验,极大地提高了模型的预测精度。卷积神经网络[8-10](Convolutional Neural Network, CNN)作为在DL学习领域中的一个关键架构,经常被研究人员应用到农业分割领域。Zhang等[11]提出了一种以UNet++为基本框架的孢子分割模型。该模型将ResNet和全连接条件随机场相结合,平均像素准确率和平均交并比分别达到97.5%和94.3%。

收稿日期:2023-01-17

基金项目:"十四五"国家重点研发计划项目(2022YFD1600602);财政部和农业农村部:国家现代农业产业技术体系资助(CARS-23-D07)

作者简介:朱轶萍,研究方向为深度学习、计算机视觉。Email:1052046559@qq.com

*通信作者:吴华瑞,博士,研究员,研究方向为农业智能系统、农业大数据智能服务。E-mail:wuhr@nercita.org.cn

Zheng等[12]设计了一种使用RGB图像的视觉算法，将芒果以实例的形式进行分割，平均精确度和平均召回率分别达到了94.7%和92.9%。王璨等[13]提出了改进的双注意力语义分割方法，能够有效识别与分割玉米，平均交并比达到94.16%，平均像素识别准确率为95.68%。刘平等[14]提出的小麦开花期判定方法，可以准确地分割识别小花和小穗，平均识别精度分别达到了91%和90.9%。Song等[15]利用基于ResNet的DeepLabV3+对猕猴桃冠层图像进行果萼、枝、丝的分割效果也能得到较高精度。CNN能够有效提取农作物信息，但主要针对背景简单且特征明显的作物，而甘蓝受土壤、光照、叶片交叠遮挡、叶球与外叶颜色纹理相似等影响，使得现有的识别模型难以对甘蓝叶球和外叶进行精细分割。相比于CNN，Transformer可以通过图像块直接的关系进行建模，自动学习图像中的注意力分割，并聚焦于最具信息量的区域[16]，能够学习到更丰富的语义信息。Zheng等[17]通过在Transformer的每一层中建模全局上下文。该编码器可以与简单的解码器相结合来提供一个强大的分割模型。Reedha等[18]通过ViT模型准确地区分作物和杂草，克服了两者因纹理相似而难以有效分割的问题。Xie等[19]提出了一种简单有效且鲁棒性强的Segformer语义分割框架，将Transformer和轻量级多层感知器解码器结合在一起。由于视觉领域大部分图像分辨率较高，直接使用Transformer进行特征提取会带来庞大的计算量，无法满足田间作物识别的实时性要求。

为了解决复杂背景下甘蓝叶球和外叶分割精度和实时性低的问题，本研究将DL引入到甘蓝叶球和外叶分割任务中，并建立真实场景下的甘蓝数据集进行实验，提出一种基于UperNet-ESA的语义分割算法。将UperNet[20]作为高效的语义分割框架，选取先进的主干网络ConvNeXt[21]，在确保分割精度的同时使用复杂度较低的模型。由于叶球和甘蓝外叶颜色纹理相似，为了进一步加强叶球区域的特征提取，在主干网络中融合了高效通道注意力机制（Efficient Channel Attention, ECA）模块。并且由于甘蓝作物形态不规则，为了加强目标边界信息的提取，将特征选择模块（Feature Selection Model, FSM）和特征对齐模块（Feature Alignment Model, FAM）集成到特征金字塔网络中来提升模型的预测结果。通过实验对比验证本研究模型具有可行性，为智能化田间甘蓝生长监测管理提供理论依据。

1 材料与方法

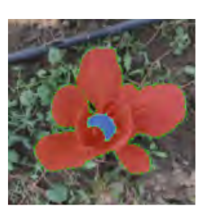
1.1 数据集构建

1.1.1 数据获取

图像数据来源于北京小汤山国家农业信息化工程技术研究中心精准农业实验基地。选用甘蓝品种为中甘-21，数据采集时间为2022年9—11月。采集的图像涵盖丰富多样的背景元素、光照差异及生长变化等真实信息，有助于提升模型在现实应用中的泛化性。采用垂直俯视角度拍摄，设备距地面高度为50 cm。图像采集的位置是室外大田，共筛选了甘蓝图像700幅。

采集的图像选用专业标注软件Labelme（v4.5.6），通过手动选择目标轮廓上的点，并将它们连接起来以绘制出符合边界形状的封闭多边形区域，自动生成保存为*.json文件。在作物叶球识别任务中，由于甘蓝叶球和外叶密切相关，需要同时标注甘蓝叶球和外叶两部分，生成掩膜。标注后的图像样本如图1所示。按照PASCAL VOC2012格式生成数据集，将数据集划分为训练集和测试集两部分，其中训练集占总数据集的80%，测试集占20%，并且两集合间无重复数据。


a. 原始图像


b. 标注

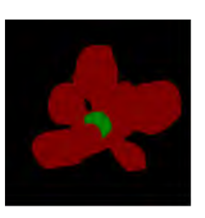

c. 标签

图1 甘蓝图像标注方法示例

Fig. 1 Example of kale image annotation method

1.1.2 数据增强

为了提升语义分割精度，通过数据增强[22]的方法来扩充训练集样本的数量，使得模型得到更为充分的训练，包括：1）亮度调整，随机选择0.35～1的值来对亮度进行调整，以此来模拟不同时间段采集的图像；2）随机旋转，以水平翻转、垂直翻转、水平垂直都翻转3种方式，来扩充不同位置得到新的图像；3）添加高斯噪声，使得模型学习图像变化的微小特征。通过上述方法随机组合将训练集扩展为原来的12倍，数据增强部分效果如图2所示。这些图像增强方法的结合模拟图像采集过程中拍摄角度和光强的变化，增加训练样本的多样性，提高模型的鲁棒性和泛化能力。



a. 原图　b. 亮度变化　c. 噪声　d. 旋转　e. 随机组合1　f. 随机组合2　g. 随机组合3　h. 随机组合4

注：e~f均表示亮度变化+旋转+噪声3种方法随机组合的数据增强图。

图2　原始及数据增强后的结球甘蓝图像

Fig. 2　Original and data enhanced images of kale

1.2 UperNet-ESA的语义分割算法

本研究主要针对田间复杂背景下甘蓝外叶和叶球的分割，通过像素值确定叶球的大小，判断甘蓝生长情况，从而进行智能化管理。语义分割作为计算机视觉的一个重要分支，人们提出了各种先进的端到端深度计算机视觉架构，包括UNet[23]、PSPNet[24]、DeepLabV3+[25]等。这些体系结构在语义分割任务上表现良好，但它们只有在特定场景处理分割任务时效果较好，本研究根据复杂背景下甘蓝叶球和外叶的形态特征，基于UperNet研究UperNet-ESA算法，关于改进的UperNet语义分割框架可以用来提高叶球识别分割精度和速度，与之前提出的几个架构不同的是，它能够在保留空间细节的同时，有效地融合不同尺度的语义信息，更适用于复杂场景下作物的特征提取。UperNet-ESA语义分割结框架如图3所示。该算法主要改进主干网络ConvNeXt，通过将ECA模块融入特征提取网络的各阶段来加强目标区域的特征提取；同时改进特征金字塔网络结构，将FAM和FSM集成到其中以加强目标边界信息的提取。

1.2.1　主干网络

主干网络负责模型的特征提取任务，考虑甘蓝叶球与外叶间存在相似性和叶片交叠等问题，使得目标轮廓定位困难，分割精度低。在计算机视觉领域，CNN一直是主流模型，但近年来由于Transformer的内在结构在各种计算机视觉任务上也表现出优异的性能。ConvNeXt充分结合CNN和Transformer的优异性能，以ResNet50作为基础，并且借鉴Swin Transformer[26]的设计理念，使得网络结构非常简洁，在提升分割精度的同时也能具有较低的模型复杂度，是目前性能较好的CNN。

对于ConvNeXt网络，根据模型计算复杂度的不同，共有4个（T/S/B/L）版本。本研究综合考虑结球甘蓝图像数据集背景复杂的特征，使用ConvNeXt-B版本。模型包括4个阶段，每个阶段主要由下采样层（Downsample）和ConvNeXt块组成，最后经全局平均池化（Global Average Pooling, GAP）、层标准化（Layer Normalization, LN）与全连接（Linear）等操作后，输出特征图，模型结构如图4a所示。为了提高模型的感受野，将每个阶段的块比设置为3∶3∶27∶3；通道数分别设为128、256、512、1 024。ConvNeXt块结构如图4b所示，首先将输入的特征图经过深度可分离卷积（Depthwise Conv2d）及LN，之后经过普通卷积（Conv2d）等操作，并且只在该模块中的1×1卷积层后加入高斯误差线性单元（Gaussian Error Linear Unit, GELU）激活函数。最后，对特征图进行一系列升维与降维处理后，使其输出原尺寸特征图。其中GELU可以被认为是ReLU的一个更平滑的变体，能有效防止过拟合。关于Downsample结构如图4c所示，主要通过一个LN加上一个卷积核大小为2，步距为2的卷积层构成。

1.2.2　注意力机制

甘蓝叶球和外叶纹理相似，在主干网络中融合注意力机制，能使网络自适应地注意甘蓝外叶和叶球目标。ECA模块具有计算效率高、模型复杂度低的优势，能够改善模型在复杂场景下特征提取的性能。本研究在ConvNeXt-B的基础上改进模型架构，将ECA集成在每一个ConvNeXt块后。通过注意权值应用于残差倒转的特征映射时，在每个块上循环

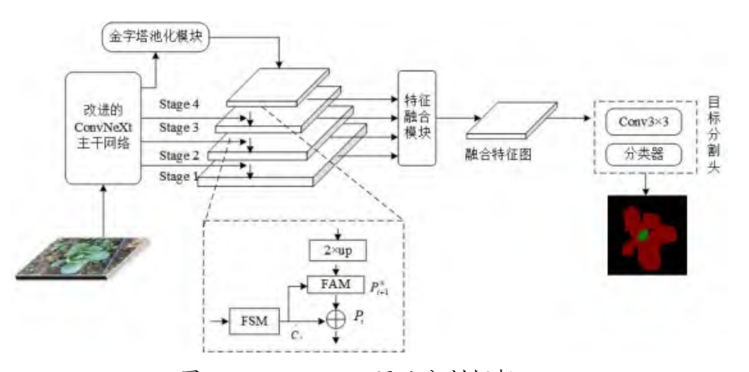


图3　UperNet-ESA语义分割框架

Fig. 3　UperNet-ESA semantic segmentation framework

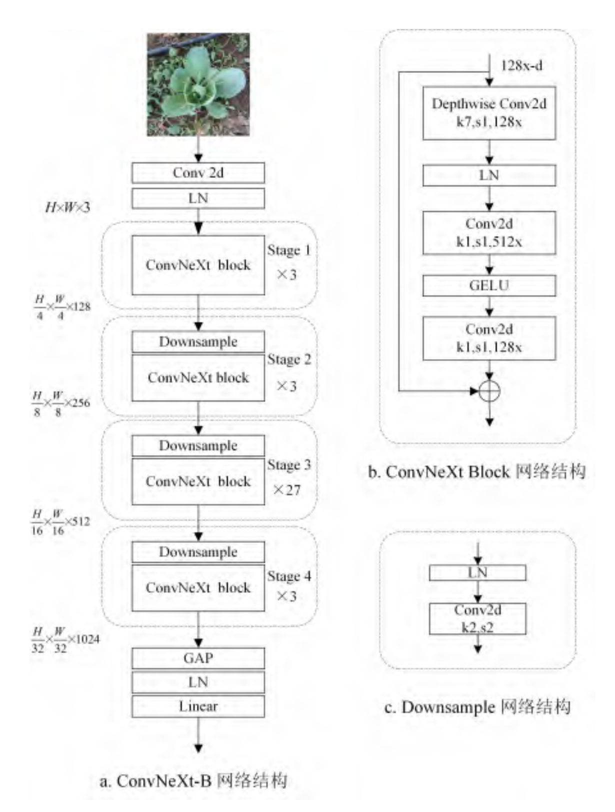


图4　ConvNeXt-B网络架构图

Fig. 4 Network architecture diagram of ConvNeXt-B

训练注意参数。然后每个块循环对象形成一个具有注意权值的特征图，实现了注意参数的重复训练，继而增强对全局特征信息的捕获。

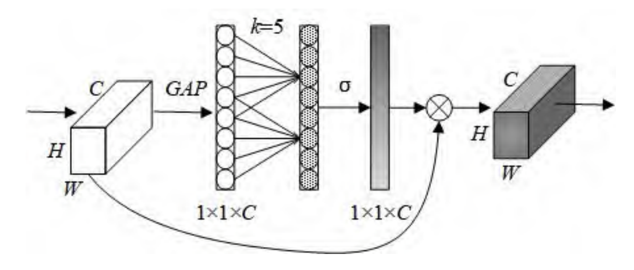
ECA模块利用GAP降低通道维度，并使用1D卷积操作对通道进行建模。通过将运算应用在通道维度上，可以更加高效地处理大规模输入数据，并减少计算和内存资源的需求。在GAP操作后确定自适应核大小k，关于k的计算如公式（1）所示；同时，利用Sigmoid激活函数（σ）进行对应权重的归一化；最后，将已生成的各个通道权重与输入特征图相加权，以获得经过通道注意力调整后的特征图。ECA模块如图5所示。

$$k = \left|\frac{\log_2(C)}{\gamma} + \frac{b}{\gamma}\right|_{odd} \tag{1}$$

式中：γ和b为函数参数，分别设为2和1；C为输入的通道数。

1.2.3　FSM和FAM模块

UperNet的模型设计主要基于金字塔池化模块和特征金字塔网络两部分组成，其中上采样和局部特征之间直接添加像素会导致具有不对齐上下文的特征映射，从而会在甘蓝叶球和外叶的边界上出现错误预测分类。由于甘蓝叶球和外叶具有纹理相似性和形状不规则等特点，更加提升模型分割难度。



注：C表示通道数；H和W分别表示高度和宽度；k表示一维卷积局部交互大小。

图5　ECA模块结构图

Fig. 5 Structure diagram of ECA module

为了解决上述问题，本研究将FSM和FAM集成到特征金字塔网络中，以此来改善模型的预测结果。

FSM能有选择地保留重要的特征映射，丢弃无用的特征映射进行信道缩减，并相应地重新校准。FSM的结构如图6所示，计算如公式（2）和公式（3）所示。

$$u = f_m(z) \tag{2}$$

$$\hat{C}_i = f_s(C_i + \boldsymbol{u} \times C_i) \tag{3}$$

式中：z为通过全局平均池化操作从输入特征映射；f_m（.）为特征重要性建模层；$\boldsymbol{u}$为特征重要性向量；C_i和$\hat{C}_i$分别为输入特征和输出特征映射；f_s（.）为特征选择层。

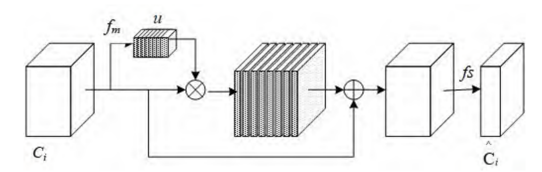


图6　特征金字塔网络的特征选择模块图

Fig. 6 Feature selection module diagram for feature pyramid networks

FAM包括一个可变形的卷积，用于学习像素的变换偏移量，以在上下文中对齐上采样的高级特征。与标准卷积不同，可变形卷积在感受野中引入了可学习的偏移量，能够更好地去覆盖目标形状周围。特征对齐的过程可以用数学方法表示，如公式（4）和公式（5）所示。

$$\Delta_i = f_\alpha([\hat{C}_{i-1}, P_i^u]) \tag{4}$$

$$\hat{P}_i^u = f_o(P_i^u, \Delta_i) \tag{5}$$

式中：P_i^u为上采样的特征图；$\hat{C}_{i-1}$为相应的自底向上特征图输入的FSM的输出。通过元素添加或通道连接的特征融合都会损害目标边界周围的预测，P_i^u与C_{i-1}之间存在可预见的空间错位。因此，在进行特征聚合前，通过$\hat{C}_{i-1}$和P_i^u的拼接，再进行

f_α（.）根据偏移量获取对齐特征$\hat{P}_i^u$。f_o（.）为从空间差中获取偏移量Δ_i的函数。FAM结构如图7所示。对于3×3的卷积层，N值为9。

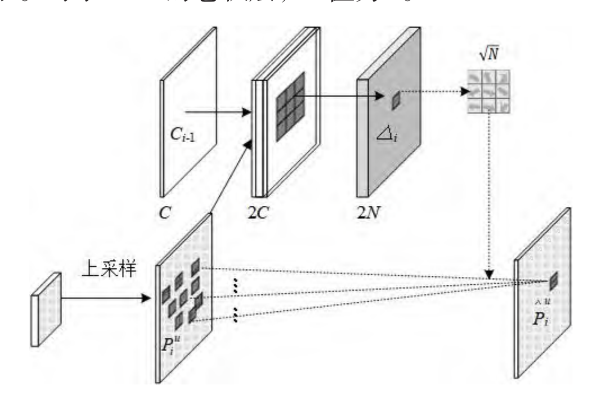


图7　特征金字塔网络的特征对齐模块图

Fig. 7 Feature alignment module diagram for feature pyramid networks

FSM自适应地聚合使用不同尺寸的卷积核、形状和方向所提取的信息，以实现精确定位；FAM通过使用学习到的偏移量调整卷积核中的每个采样位置来学习将上采样特征映射对齐到一组参考特征映射；将FSM和FAM这两个模块集成在一个自上而下的特征金字塔网络中，以实现目标边界框的精准预测。

2　实验与分析

2.1 实验平台配置

本研究实验环境基于Window 10操作系统的计算机，中央处理器（CPU）为AVX2，图形处理器（GPU）为 NVIDIA GeForce RTX 2080 Ti。编程语言为python 3.8，通用并行计算架构为CUDA 10.2，cuDNN 7.6.5用于深度神经网络的GPU加速库，计算机视觉库为OpenCV 4.7.0，采用Pytorch 1.7.1深度学习框架对本研究分割模型进行搭建和参数调整。

2.2 评价指标

本研究模型的性能评价指标主要采用平均像素准确率（Mean Pixel Accuracy, mPA）、平均交并比（Mean Intersection over Union, mIoU），以及模型的推理速度fps来评估图像分割模型性能。其中mPA衡量了模型对各个类别的像素进行正确预测的平均准确率，越高的mPA值表示模型的像素预测精度越好，计算如公式（6）所示。mIoU通过计算预测分割结果与真实分割结果的交集和并集之比来衡量模型的分割准确度。该指标越高表示预测结果与真实结果的重合度越高，模型的分割效果也就越好。定义如公式（7）所示。

$$\mathrm{mPA} = \frac{1}{N+1}\sum_{i=0}^{N}\frac{n_{ii}}{t_i} \tag{6}$$

$$\mathrm{mIoU} = \frac{1}{N+1}\sum_{i=0}^{N}\frac{n_{ii}}{t_i + \sum_{j=0}^{N} n_{ji} - n_{ii}} \tag{7}$$

式中：N为分割的目标类别数量（无背景的情况），个；n_{ii}为正确分类的像素数，PPI；$t_i = \sum_{i=0}^{N} n_{ij}$为目标类别$i$的像素数，PPI；$n_{ij}$为目标类别$i$被预测为类别$j$的像素数，PPI；$n_{ji}$为目标类别$j$被预测为类别$i$的像素数，PPI。

2.3 结果分析

2.3.1　网络训练

将训练效果最好的文件保存为权重文件，然后用于测试。为了提高模型的训练效果，采用Adam优化器，设置初始学习率为5e−4，权重衰减为0.01，动量因子为0.9。在训练过程中，共进行300轮的迭代。其中，前50轮被用于进行冻结训练，在最初训练过程中，可能会对模型的某些部分（如权重）进行限制，以便稳定模型的初始学习。然后，在接下来的250轮训练中，对整个模型进行解冻，允许所有参数进行更新和优化。这种冻结和解冻的训练策略有助于模型在初始阶段进行有针对性的学习，然后再通过更长时间的训练来进一步提升性能。采用交叉熵损失函数（Loss）衡量真实值与模型预测值间的差异，计算如公式（8）所示。

$$\mathrm{Loss} = -\frac{1}{M}\sum_{i=1}^{M}\sum_{n=1}^{N} y_i^N \ln(\hat{y}_i^N) \tag{8}$$

式中：M为单一批量中像素数的总和，PPI；N为目标分类数，个；y_i^N为像素点i对类别的标注，正确类时该值为1，否则为0；$\hat{y}_i^N$为像素点i为类别N的概率值。

在模型学习过程中，记录每个Epoch的训练损失值和验证损失值，并利用Origin绘制如图8所示的损失值变化曲线。由图8可知，模型在训练初期经历快速的学习和调整阶段，在达到约200次迭代之后，模型的性能变化变得相对较小，可以认为模型趋于收敛状态。

2.3.2　不同主干网络的性能对比

为了验证本研究的基础模型对甘蓝研究是最优架构，采用相同的甘蓝数据集和语义分割框架UperNet，选用性能较好的主干网络ResNet-50、Swin

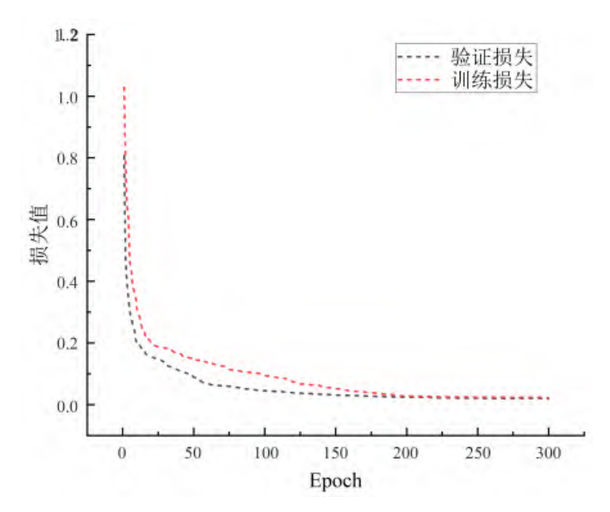


图8　UperNet-ESA模型损失值变化曲线

Fig. 8 The loss value curve of UperNet-ESA model

Transformer和ConvNeXt进行训练对比，结果如表1所示。

表1　基于UperNet框架的不同主干网络的性能对比

Table 1 Performance comparison of different backbone networks based on the UperNet framework

主干网络	mPA/%	mIoU/%	分割速度/(帧/s)
ResNet-50	88.86	88.72	17.6
Swin Transformer	89.81	89.19	15.5
ConvNeXt	90.17	90.12	17.4

由表1可知，ConvNeXt作为主干网络在测试集上的表现较好，预测精度最高。其中mPA可达到90.17%，比ResNet-50和Swin Transformer分别提高1.31%和0.36%。mIoU为90.12%，相较于ResNet-50和Swin Transformer分别提高1.4%和0.93%。在分割速度方面，主干网络ConvNeXt的处理速度与ResNet-50的差不多，但是精度明显比ResNet-50高。通常情况下，要实现精确的像素级语义分割，网络结构更为复杂，分割速度会有所下降。ConvNeXt网络将多个不同尺寸的卷积核组合成一个大的卷积核。这种方法比传统的卷积核更加灵活，能够捕获更多的局部特征。此外，ConvNeXt还加入自注意力机制，原理与Swin Transformer类似，且采用分组卷积，进一步降低参数数量和计算复杂度。所以ConvNeXt整体性能处于较好的水平，因此本研究选择的基础架构模型以UperNet为高效语义分割框架，ConvNeXt作为主干网络。

2.3.3　消融实验

采用消融实验来验证模型中不同部位改进的有效性。针对ECA模块、FAM和FSM模块，设计3组实验，并在测试集上使用mIoU和mPA进行评估。模型1以原始ConvNeXt-B作为主干网络，UPerNet作为语义分割框架作为模型的基础架构。模型2将ECA模块融合到模型1的ConvNeXt-B中。最后，本研究UperNet-ESA为模型3，是在模型2的基础上将FAM和FSM模块集成UperNet框架中。3组实验对比结果如表2所示。

表2　基于UperNet−ConvNeXt的消融实验

Table 2 UperNet-ConvNeXt based ablation experiments

模型	ECA	FAM+FSM	mPA/%	mIoU/%
模型1	×	×	90.17	90.12
模型2	√	×	92.88	90.14
模型3	√	√	94.32	92.45

由表2可以看出，本研究改进的模型3表现最好，mPA和mIoU分别达到94.32%和92.45%。通过对比模型1和模型2可以看出，将ECA集成到主干网络ConvNeXt-B中有助于提高模型的语义分割性能，mPA和mIoU分别提高2.71%和0.02%。通过对比模型3和模型2的实验结果可以发现，在模型的UperNet框架中加入FAM和FSM模块后，mPA和mIoU分别增加1.44%和2.31%。以上3个消融实验充分验证了ECA、FAM和FSM的有效性和合理性。

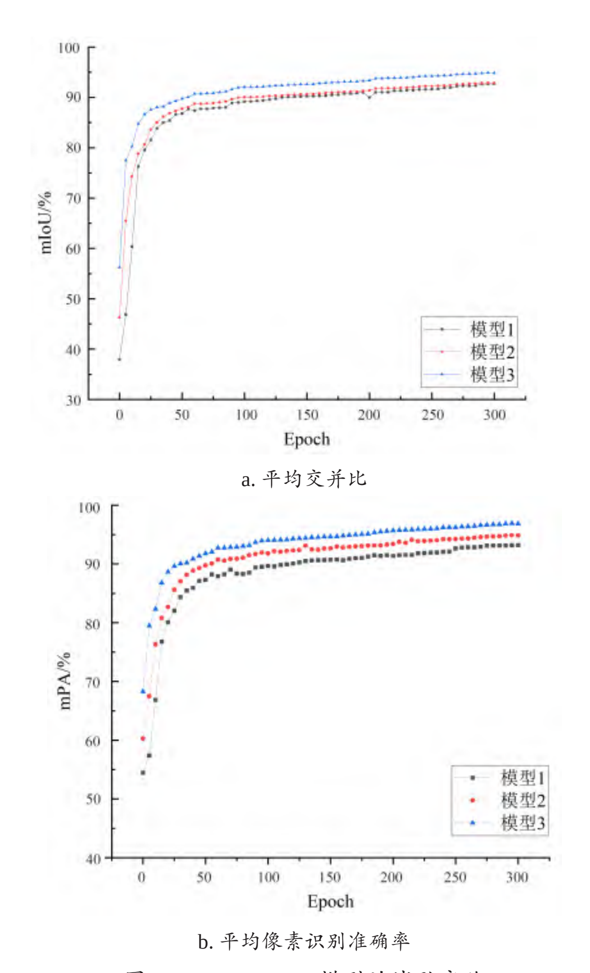
在训练过程中，为了考察各网络架构分割性能的变化情况，每5次迭代进行一次评估，mIoU和mPA的变化结果如图9所示。总体来看，各模型的mIoU和mPA都随着epoch的增加而增加，最终趋于收敛。本研究甘蓝叶球分割模型的mIoU在整个训练中均高于其他网络架构。其中模型1和模型2的mIoU曲线相接近，模型3由于FAM和FAM模块的融合，极大提升了目标边界的分割精度。

综上述分析可知，结球甘蓝叶球语义分割的性能在整体训练过程中，优于原始网络，且能够通过学习获得更大的提升。说明相关的改进对模型的分割性能和像素预测精有一定程度的提高。

2.3.4　不同网络架构的性能对比

为了验证本研究提出的UperNet-ESA算法在田间甘蓝数据集上分割的优势，与当前主流的UNet模型、PSPNet模型、以ResNet-50、MobilenetV2、Xception为主干的DeepLabV3+模型进行比较。从mPA、mIoU、FPS这3个指标进行总体分析。本研究改进的模型相对其他模型的各个评价指标如表3所示。

由表3结果可知，本研究提出的UperNet-ESA模型在分割精度方面优于其他模型。本研究的mPA



a. 平均交并比

b. 平均像素识别准确率

图9 UperNet-ESA模型的消融实验

Fig. 9 Ablation experiments with the UperNet-ESA model

表3 UperNet-ESA研究不同网络架构的性能对比

Table 3 Performance comparison of different network architectures for the UperNet-ESA study

网络	主干网络	mPA/%	mIoU/%	分割速度/(帧/s)
UNet	ResNet-50	82.80	80.14	15.2
PSPNet	ResNet-50	80.76	79.41	20.3
DeeplabV3+	ResNet-50	85.64	81.80	16.2
DeeplabV3+	MobilenetV2	90.01	89.19	14.4
DeeplabV3+	Xception	88.11	85.34	15.7
UperNet +FAM+FSM	ConvNeXt-B +ECA	94.32	92.45	16.6

为94.32%，比UNet模型、PSPNet模型、以ResNet-50、MobilenetV2、Xception为主干的DeepLabV3+模型的mPA值分别提高11.52%、13.56%、8.68%、4.31%和6.21%。同样的，本研究的mIoU为92.45%，比基于UNet模型、PSPNet模型、以ResNet-50、MobilenetV2、Xception为主干的DeepLabV3+模型的mIoU分别提高12.21%、13.04%、10.65%、3.26%和7.11%。主要原因是本模型引入ECA模块，以及对特征金字塔网络的改进，强化各阶段对目标特征的判断，获取有效的全局上下文信息。此外，虽然PSPNet模型的推理速度最快，但是整体精度太低，不适用于构建甘蓝语义分割模型。本研究模型推理速度比其他网络模型的推理速度都要快。综合分析，本研究模型能够在保证模型准确率的前提下尽可能花较少的推理时间，以实现甘蓝叶球识别精度与识别速度的平衡，从而为智能化田间管理提供理论依据。

2.3.5 可视化分析

为了验证本研究方法在田间甘蓝图像分割任务中的有效性和可解释性，对测试集中的图像进行推理，并进行可视化分析。由图10所示，每一行分别对应甘蓝未结球、甘蓝初结球、甘蓝结球紧实图，每一列分别对应原图、人工分割图、UNet算法分割图、PSPNet算法分割图、DeepLabV3+（ResNet）算法分割图、DeepLabV3+（Mobile-netV2）算法分割图、DeepLabV3+（Xception）算法分割图和本研究算法分割图。

由图10可得，本研究算法对于复杂田间场景中的甘蓝叶球和外叶分割具有较好的性能，分割效果同真实值间无明显差异。错误分割通常在目标边缘附近发生。只有少数像素被错误地分类。这对于整体分割效果来说影响较小，人眼很难检测到这些细微差异。由图10c（3）和图10d（3）可以看出，在甘蓝外叶相对较大的情况下，对叶球容易产生错误分割。由图10c（1）和图10d（3）可知，在甘蓝未结球的情况下也会出现叶球识别错误的情况。因而可以得出结论，UNet、PSPNet作为图像分割经典模型，在叶球分割识别上准确率较低，不适用于甘蓝作物图像分割算法。对于图10e（2）、图10f（2）、图10e（3）和图10f（3）分析可得，算法对甘蓝外叶间的间隙分割不明确，在植物叶片分割上存在轮廓不清晰的问题。相对而言，Deep-labV3+（Mobilenet）模型的分割性能与本研究模型最为接近，但它没有充分强调各个阶段的判别特征。这可能导致同一类别内的预测结果不一致。结合表3综合可得，本研究方法在甘蓝不同生长状态下的像素识别精确度上具有更强的实际泛化能力。此外，本研究的算法简洁高效，具有较强的实时性。

3 结 论

大田复杂背景下甘蓝叶球的精准快速分割识别利于智能化管理，而甘蓝作物通常具有大小不一、

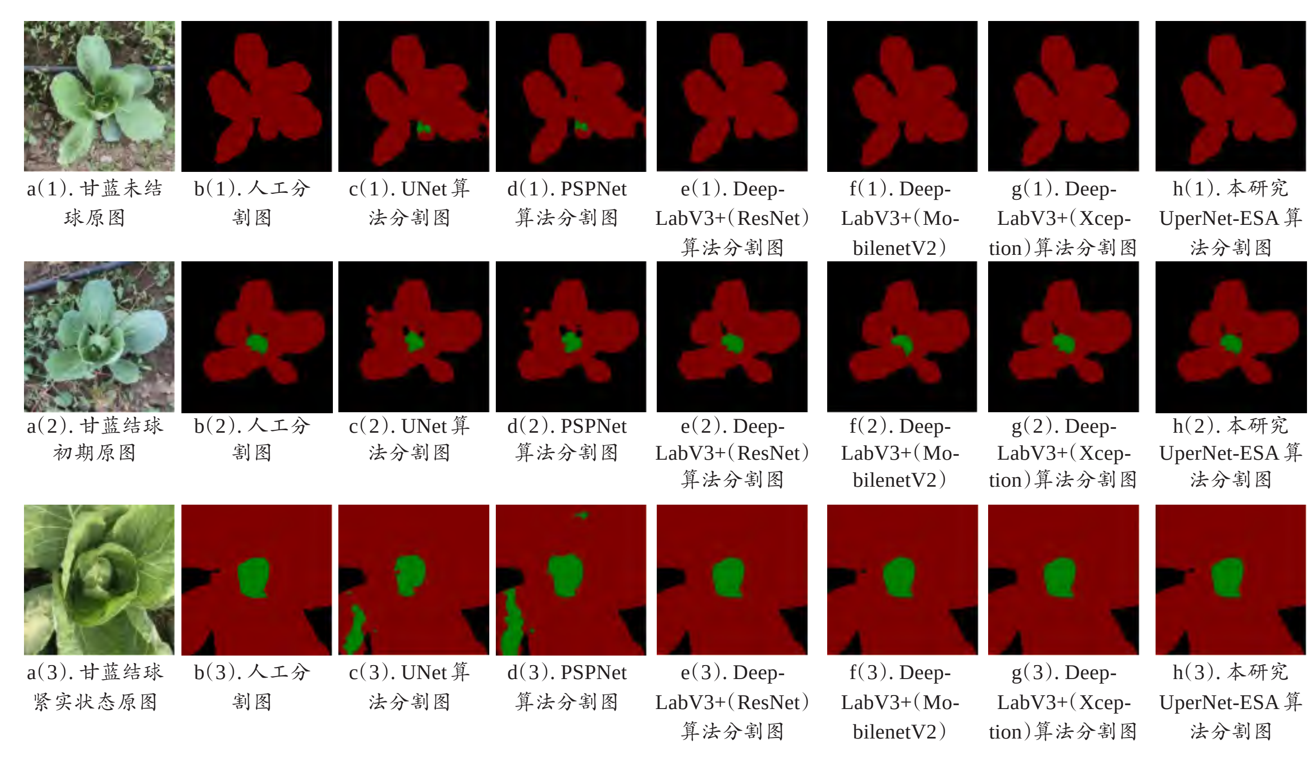


图10　UperNet-ESA模型与其他网络模型的分割效果对比图

Fig. 10　Comparison of segmentation effect of UperNet-ESA model with other network models

光线差异、形状不规则、叶片交叠遮挡和目标纹理相似等特点，这对分割模型提出了更高的要求。本研究针对性地选择更适应于复杂背景下解决作物分割问题的高效语义分割框架UperNet建立UperNet-ESA算法，将先进的ConvNeXt作为主干网络。通过自建甘蓝数据集进行实验分析，根据甘蓝作物的特点，再进行基础网络的改进。主要改进包括主干网络和ECA模块的融合、特征金字塔网络和FAM、FSM的集成以及网络训练参数的调整。

实验结果表明，本研究提出的UperNet-ESA模型在性能上优于原始网络；且与当前主流的语义分割网络相比，本研究的改进模型能够达到最佳精度-速度平衡效果，其中mIoU和mPA分别达到92.45%和94.32%，推理速度可达16.6 fps。在后续研究中，将对现有的模型进一步优化和改进，同时丰富甘蓝数据集，以包含更多不同种类的结球甘蓝叶球样本，为智能化甘蓝田间管理提供更为可靠和全面的理论支持。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] 岳智臣, 俞国红, 薛向磊, 等. 杭州秋季露地甘蓝轻简化增效栽培技术简析[J]. 浙江农业科学, 2023, 64(5): 1103-1106.
YUE Z C, YU G H, XUE X L, et al. Analysis of light and simple and efficient cultivation techniques of autumn cabbage in Hangzhou[J]. Journal of Zhejiang agricultural sciences, 2023, 64(5): 1103-1106.

[2] 陈皓颖. 人工智能在农业领域中的应用[J]. 灌溉排水学报, 2023, 42(7): 146.
CHEN H Y. Application of artificial intelligence in agricultural field[J]. Journal of irrigation and drainage, 2023, 42(7): 146.

[3] 刘海桥, 刘萌, 龚子超, 等. 基于深度学习的图像匹配方法综述[J/OL]. 航空学报, (2024-01-16). https://kns.cnki.net/kcms/detail/11.1929.V.20240115.1139.004.html.
LIU H Q, LIU M, GONG Z C, et al. A review of image matching methods based on deep learning[J/OL]. Acta aeronautica et astronautica sinica, (2024-01-16). https://kns. cnki. net/kcms/detail/11.1929. V. 20240115.1139.004.html.

[4] 赵永强, 金芝, 张峰, 等. 深度学习图像描述方法分析与展望[J]. 中国图象图形学报, 2023, 28(9): 2788-2816.
ZHAO Y Q, JIN Z, ZHANG F, et al. Deep-learning-based image captioning: Analysis and prospects[J]. Journal of image and graphics, 2023, 28(9): 2788-2816.

[5] MINAEE S, BOYKOV Y, PORIKLI F, et al. Image segmentation using deep learning: A survey[J]. IEEE trans pattern anal mach intell, 2022, 44(7): 3523-3542.

[6] GAN P X, LUO X Y, LIU B, et al. Research on semantic segmentation method of urban streetscape image based on deep learning[C]// Seventh Asia Pacific Conference on Optics Manufacture and 2021 International Forum of Young Scientists on Advanced Optical Manufacturing (APCOM and YSAOM 2021). Burlingame, California,

USA: SPIE, 2022.

[7] 翁杨, 曾睿, 吴陈铭, 等. 基于深度学习的农业植物表型研究综述[J]. 中国科学(生命科学), 2019, 49(6): 698-716.
WENG Y, ZENG R, WU C M, et al. A survey on deep-learning-based plant phenotype research in agriculture[J]. Scientia sinica (vitae), 2019, 49(6): 698-716.

[8] 刘俊奇, 涂文轩, 祝恩. 图卷积神经网络综述[J]. 计算机工程与科学, 2023, 45(8): 1472-1481.
LIU J Q, TU W X, ZHU E. Survey on graph convolutional neural network[J]. Computer engineering & science, 2023, 45(8): 1472-1481.

[9] 郭庆梅, 于恒力, 王中训, 等. 基于卷积神经网络的图像分类模型综述[J]. 电子技术应用, 2023, 49(9): 31-38.
GUO Q M, YU H L, WANG Z X, et al. Review of image classification models based on convolutional neural networks[J]. Application of electronic technique, 2023, 49 (9): 31-38.

[10] 张鑫, 姚庆安, 赵健, 等. 全卷积神经网络图像语义分割方法综述[J]. 计算机工程与应用, 2022, 58(8): 45-57.
ZHANG X, YAO Q A, ZHAO J, et al. Image semantic segmentation based on fully convolutional neural network[J]. Computer engineering and applications, 2022, 58 (8): 45-57.

[11] ZHANG D Y, ZHANG W H, CHENG T, et al. Segmentation of wheat scab fungus spores based on CRF_ResUNet++[J]. Computers and electronics in agriculture, 2024, 216: ID 108547.

[12] ZHENG C, CHEN P F, PANG J, et al. A mango picking vision algorithm on instance segmentation and key point detection from RGB images in an open orchard[J]. Biosystems engineering, 2021, 206(6): 32-54.

[13] 王璨, 武新慧, 张燕青, 等. 基于双注意力语义分割网络的田间苗期玉米识别与分割[J]. 农业工程学报, 2021, 37 (9): 211-221.
WANG C, WU X H, ZHANG Y Q, et al. Recognition and segmentation of maize seedlings in field based on dual attention semantic segmentation network[J]. Transactions of the Chinese society of agricultural engineering, 2021, 37 (9): 211-221.

[14] 刘平, 刘立鹏, 王春颖, 等. 基于机器视觉的田间小麦开花期判定方法[J]. 农业机械学报, 2022, 53(3): 251-258.
LIU P, LIU L P, WANG C Y, et al. Determination method of field wheat flowering period baesd on machine vision[J]. Transactions of the Chinese society for agricultural machinery, 2022, 53(3): 251-258.

[15] SONG Z Z, ZHOU Z X, WANG W Q, et al. Canopy segmentation and wire reconstruction for kiwifruit robotic harvesting[J]. Computers and electronics in agriculture, 2021, 181: ID 105933.

[16] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. arXiv: 2010.11929, 2020.

[17] ZHENG S X, LU J C, ZHAO H S, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[EB/OL]. arXiv: 2012.15840, 2020.

[18] REEDHA R, DERICQUEBOURG E, CANALS R, et al. Transformer neural network for weed and crop classification of high resolution UAV images[J]. Remote sensing, 2022, 14(3): ID 592.

[19] XIE E Z, WANG W H, YU Z D, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[J]. arXiv: 2105.1520, 2021.

[20] XIAO T T, LIU Y C, ZHOU B L, et al. Unified perceptual parsing for scene understanding[M]// Computer Vision – ECCV 2018. Cham: Springer International Publishing, 2018: 432-448.

[21] LIU Z, MAO H, WU C Y, et al. A ConvNet for the 2020s[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, Louisiana, USA: IEEE, 2020: 11966-11976.

[22] TAKAHASHI R, MATSUBARA T, UEHARA K. Data augmentation using random image cropping and patching for deep CNNs[J]. IEEE transactions on circuits and systems for video technology, 2020, 30(9): 2917-2931.

[23] DIAO Z H, GUO P L, ZHANG B H, et al. Maize crop row recognition algorithm based on improved UNet network[J]. Computers and electronics in agriculture, 2023, 210: ID 107940.

[24] YANG C Z, GUO H J. A method of image semantic segmentation based on PSPNet[J]. Mathematical problems in engineering, 2022, 2022: ID 8958154.

[25] 马冬梅, 李鹏辉, 黄欣悦, 等. 改进DeepLabV3+的高效语义分割[J]. 计算机工程与科学, 2022, 44(4): 737-745.
MA D M, LI P H, HUANG X Y, et al. Efficient semantic segmentation based on improved DeepLabV3+ [J]. Computer engineering & science, 2022, 44(4): 737-745.

[26] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2021: 10012-10022.

Identification Method of Kale Leaf Ball Based on Improved UperNet

ZHU Yiping[1,2], WU Huarui[1,2,3,4*], GUO Wang[2,3,4], WU Xiaoyan[2]

(1. *School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang* 212013, *China*; 2. *National Engineering Research Center for Information Technology in Agriculture, Beijing* 100097, *China*; 3. *Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing* 100097, *China*; 4. *Key Laboratory of Digital Village Technology, Ministry of Agriculture and Rural Affairs, Beijing* 100097, *China*)

Abstract:

[Objective] Kale is an important bulk vegetable crop worldwide, its main growth characteristics are outer leaves and leaf bulbs. The traits of leaf bulb kale are crucial for adjusting water and fertilizer parameters in the field to achieve maximum yield. However, various factors such as soil quality, light exposure, leaf overlap, and shading can affect the growth of in practical field conditions. The similarity in color and texture between leaf bulbs and outer leaves complicates the segmentation process for existing recognition models. In this paper, the segmentation of kale outer leaves and leaf bulbs in complex field background was proposed, using pixel values to determine leaf bulb size for intelligent field management. A semantic segmentation algorithm, UperNet-ESA was proposed to efficiently and accurately segment nodular kale outer leaf and leaf bulb in field scenes using the morphological features of the leaf bulbs and outer leaves of nodular kale to realize the intelligent management of nodular kale in the field.

[Methods] The UperNet-ESA semantic segmentation algorithm, which uses the unified perceptual parsing network (UperNet) as an efficient semantic segmentation framework, is more suitable for extracting crop features in complex environments by integrating semantic information across different scales. The backbone network was improved using ConvNeXt, which is responsible for feature extraction in the model. The similarity between kale leaf bulbs and outer leaves, along with issues of leaf overlap affecting accurate target contour localization, posed challenges for the baseline network, leading to low accuracy. ConvNeXt effectively combines the strengths of convolutional neural networks (CNN) and Transformers, using design principles from Swin Transformer and building upon ResNet50 to create a highly effective network structure. The simplicity of the ConvNeXt design not only enhances segmentation accuracy with minimal model complexity, but also positions it as a top performer among CNN architectures. In this study, the ConvNeXt-B version was chosen based on considerations of computational complexity and the background characteristics of the knotweed kale image dataset. To enhance the model's perceptual acuity, block ratios for each stage were set at 3:3:27:3, with corresponding channel numbers of 128, 256, 512 and 1 024, respectively. Given the visual similarity between kale leaf bulbs and outer leaves, a high-efficiency channel attention mechanism was integrated into the backbone network to improve feature extraction in the leaf bulb region. By incorporating attention weights into feature mapping through residual inversion, attention parameters were cyclically trained within each block, resulting in feature maps with attentional weights. This iterative process facilitated the repeated training of attentional parameters and enhanced the capture of global feature information. To address challenges arising from direct pixel addition between up-sampling and local features, potentially leading to misaligned context in feature maps and erroneous classifications at kale leaf boundaries, a feature alignment module and feature selection module were introduced into the feature pyramid network to refine target boundary information extraction and enhance model segmentation accuracy.

[Results and Discussions] The UperNet-ESA semantic segmentation model outperforms the current mainstream UNet model, PSPNet model, DeepLabV3+ model in terms of segmentation accuracy, where mIoU and mPA reached 92.45% and 94.32%, respectively, and the inference speed of up to 16.6 frames per second (fps). The mPA values were better than that of the UNet model, PSPNet model, ResNet-50 based, MobilenetV2, and DeepLabV3+ model with Xception as the backbone, showing improvements of 11.52%, 13.56%, 8.68%, 4.31%, and 6.21%, respectively. Similarly, the mIoU exhibited improvements of 12.21%, 13.04%, 10.65%, 3.26% and 7.11% compared to the mIoU of the UNet-based model, PSPNet model, and DeepLabV3+ model based on the ResNet-50, MobilenetV2, and Xception backbones, respectively. This performance enhancement can be attributed to the introduction of the ECA module and the improvement made to the feature pyramid network in this model, which strengthen the judgement of the target features at each stage to obtain effective global contextual information. In addition, although the PSPNet model had the fastest inference speed, the overall accuracy was too low to for developing kale semantic segmentation models. On the contrary, the proposed model exhibited superior inference speed compared to all other network models.

[Conclusions] The experimental results showed that the UperNet-ESA semantic segmentation model proposed in this study outperforms the original network in terms of performance. The improved model achieves the best accuracy-speed balance compared to the current mainstream semantic segmentation networks. In the upcoming research, the current model will be further optimized and enhanced, while the kale dataset will be expanded to include a wider range of samples of nodulated kale leaf bulbs. This expansion is intended to provide a more robust and comprehensive theoretical foundation for intelligent kale field management.

Key words: kale; semantic segmentation; leafball identification; attention mechanism; feature selection; feature alignment

Foundation items: National Key Research and Development Programme (2022YFD1600602); Ministry of Finance and Ministry of Agriculture and Rural Development: Funding for the National Modern Agricultural Industry Technology System (CARS-23-D07)

Biography: ZHU Yiping, E-mail: 1052046559@qq.com

*__Corresponding author:__ WU Huarui, E-mail: wuhr@nercita.org.cn